

## Deferred automatic disclosure: Ensuring data access and protecting the “right to first use”.<sup>1</sup>

Giovanni Capoccia (University of Oxford)

[Prepared for the Newsletter of the International History and Politics Section, APSA, 2/2016]

After an initial period in which the debate on the appropriateness of the DA-RT guidelines for political science research rightly focused on issues of sensitivity, confidentiality and burden of data disclosure, relevant for various forms of qualitative research, the issue of properly ensuring the “right of first use” of original datasets is now rightly becoming an important additional topic for discussion. The “right to first use” refers to the possibility of an embargo period for original data after the first publication of results, so that the researcher can make full use of the data for other publications before releasing them to the public. Although in principle this is an issue for all types of newly collected evidence, it is particularly important for original *quantitative* datasets, which are often more immediately usable by others than qualitative data.<sup>2</sup>

Despite being included in the 2012 APSA Ethics guidelines,<sup>3</sup> this issue has been relatively neglected in the DA-RT debate. This, however, is rapidly changing. The response by six distinguished colleagues to the decision of APSA to move forward with DA-RT despite a popular petition to delay its implementation mentions that the DA-RT guidelines endanger “...the ability to publish out of original data sets without being required to share them too early”<sup>4</sup>. The letter of 7 January 2016 by 20 past, present, and future APSA Presidents openly defines the JETS statement as “...flawed in neglecting the admonition in the APSA Ethics Guidelines that the creators of new data sets should have first access and a period of personal use before making them available”.<sup>5</sup>

The criticism of the JETS standards as “flawed” may come across as direct, but it has significant merit. Of the twenty-seven journals that subscribed to the JETS guidelines, at the time of writing (January 2016) seventeen do not mention at all on their website the possibility of embargoes for original data in their “Instructions for Authors” or similar documentation. (In one case, the possibility of an embargo to protect the “right to full use” is explicitly ruled out.) What is more, the other ten journals generally qualify the possibility of embargoing original data as an “exemption” from the rule of immediate disclosure, not as a right. Authors can apply to the editors for an exemption but the editors retain full

---

<sup>1</sup> I thank Steve Hanson for comments on a previous version of this piece.

<sup>2</sup> Most of what I say about data access and right to first use applies to original qualitative data too. These, of course, present distinctive and important problems, including for example the often excessive burden of data disclosure that the imposition of DA-RT guidelines would entail for researchers. For the purpose of this short contribution, I focus on quantitative data.

<sup>3</sup> <http://www.dartstatement.org/#!2012-apsa-ethics-guide-changes/c13ay> Point 6.6. Accessed January 25, 2016.

<sup>4</sup> <http://dialogueondart.org/2015/11/25/response-apsa-leadership-risks-pre-emptying-necessary-deliberation/> Accessed January 25, 2016.

<sup>5</sup> <http://www.politicalsciencenow.com/letter-from-distinguished-political-scientists-urging-nuanced-journal-interpretation-of-jets-policy-guidelines/> Accessed January 25, 2016.

discretion over whether it will be granted. No criteria are spelled out for granting the exemption and therefore there is no way for authors to know whether they can reasonably expect such an exemption.

The language of “exemptions” hardly accords with treating the “right of first use” as a *right*, which, by definition, should at the very least give the researcher clear guarantees in advance that her data will be protected for long enough to allow her to make full use of them. Importantly, the failure to clearly protect first use disincentivizes the collection of original and innovative data to answer new questions. This has already been noted by other colleagues<sup>6</sup>, but it is perhaps worthwhile briefly restating the point here. The position of the JETS journals outlined above risks creating powerful disincentives for scholars – in particular for individual scholars-- to invest in the creation of innovative datasets. Many would reasonably think that the large amount of time and resources involved would not be adequately rewarded: simply being cited as the author of a new dataset would not, in most cases, be seen as a sufficient incentive. Graduate students in particular would have no incentives to embark in the construction of innovative datasets. The option of analyzing “off-the-shelf” datasets, which of course have no cost of collection and no problems of disclosure, would look more attractive to many than embarking in the expensive operation of generating innovative datasets. This in turn is likely to impact on the type of questions that researchers ask: the prevailing incentive would be to ask questions that can be answered with *existing* data, rather than to invest substantive resources to collect original data to answer new questions, for which adequate empirical data may not exist.

#### **“Deferred automatic disclosure”: Decoupling the release of data to the journal from release to the wider public**

The key objection to providing clear criteria for embargoing new data is that, if the author is not required to disclose data at the time of publication, there is no way to force data disclosure at the end of the embargo period. There is no guarantee that the data used in the study will ever be made accessible and therefore it must imposed as a condition of publication.<sup>7</sup>

This is an important concern, and it should be taken seriously. One response would be to decouple the release to the journal of the original data from their release to the wider public. This could be implemented in various ways, but the basic principle is simply stated: An author using original data would be required to release the data *to the journal*, or to a public repository such as Dataverse or ICPSR, as a condition of publication. The journal, or the repository, would in turn not release the data *to the wider public* before an agreed upon date. Such a “deferred automatic disclosure” system would not only provide the infrastructure for guaranteeing an adequate embargo for original data but seems to present several advantages over the current situation: First, it would solve the moral hazard problem mentioned earlier and *guarantee* data access for the wider community, albeit only at the end of an

---

<sup>6</sup> See <http://dialogueondart.org/2015/11/11/bryan-jones-on-how-da-rt-may-undermine-the-production-of-collective-goods/> and <https://www.washingtonpost.com/news/monkey-cage/wp/2015/11/09/political-scientists-are-debating-a-new-initiative-to-make-research-more-trustworthy-heres-why-im-skeptical/> Accessed January 25, 2016.

<sup>7</sup> See e.g. Savage CJ, Vickers AJ (2009) Empirical Study of Data Sharing by Authors Publishing in PLoS Journals. PLoS ONE 4(9): e7078. doi:10.1371/journal.pone.0007078

embargo period. Second, it would provide *more* transparency and data access than the current system because the data and coding operations at the basis of *all* published quantitative studies (and not only the articles not exempted by the editors) would be eventually released to the public. Third, transparency for authors would also be enhanced because the decision to release their data would no longer depend on “exemptions” granted by individual journal editors based on unstated criteria. Fourth, a system that guarantees deferred disclosure would allow a more open discipline-wide discussion on the general criteria to be used to establish the length of an embargo. Clearly an excessive or arbitrary length would be detrimental to data access, so any embargo should certainly be limited in time. At the same time though, the length of the embargo should be adequate to protect the “right to first use” (following principles similar to those of copyright and patent law) by giving the researcher a reasonable amount of time to make full use of the data. The APSA Ethics guidelines state that embargoes should last one year (point 6.6.). This might be appropriate in some cases, but not necessarily in all cases. In some cases, the nature of the data and the professional condition of the researcher (e.g. in terms of teaching load and available resources) might require a longer embargo period.

At the same time, a system of deferred automatic disclosure would retain some key mechanisms necessary for transparency. If the journal’s policy is to replicate the results in-house before agreeing to publish the article (as is the case for some journals, e.g. AJPS), it could still do so. Moreover, the system would still provide full transparency on the exact date of public release of the data, which, once agreed upon between the editors and the author based on the discipline’s guidelines, could be clearly stated in the first footnote of the article.

Depending on how it is implemented, a system of “deferred automatic disclosure” such as the one outlined above may impose small costs on journals or on repositories. They would have to release the data by the agreed end-of-embargo date and would have to commit to protecting the data from outside access until that time. Some journals may not have the internal resources to monitor the release of data one or more years after publication of the article. Some attention to logistics would be needed when journals change editors to reliably transfer embargoed data from the old to the new editorial team. These, however, do not appear to be insurmountable problems. For example, a system could be put in place by which, once the data are lodged with the journal as a condition of publication, the journal could immediately transfer the data to a public repository and destroy its copy.<sup>8</sup> The repository could in turn put in place a system of “automatic release” to the public after a certain date. Individual journals, of course, could still opt out of the system altogether if they wished.

Other issues of detail will certainly emerge in the discussion, and it is beyond the scope of this short piece to address them all. What is key here is that a system based on the principle of decoupling

---

<sup>8</sup> Similar practices are already current in some journals: for example, *State Politics and Policy Quarterly* allows for a limited embargo not for the data used in an analysis but “for any additional information included in the source dataset”. The author should still release this information to the journal, however. The journal editorial staff “...guarantee that the embargoed data will be used strictly to verify the integrity of the replication materials, and will not be retained after that has been established”. See [http://spa.sagepub.com/site/misc/SPPQ\\_Instructions/guidelines\\_for\\_preparing\\_replication\\_files.pdf](http://spa.sagepub.com/site/misc/SPPQ_Instructions/guidelines_for_preparing_replication_files.pdf) Accessed January 5, 2016.

disclosure to the journal from disclosure to the wider public (however implemented in detail) would provide a better framework for protecting adequately the “right to first use” than under the current scenario. At the same time, it would guarantee data access, even though not immediately for newly collected data, and indeed improve it, since would no longer allow wholesale “exemptions” from data release. To reiterate, adequately protecting the “right to first use” of original data is important not only because it constitutes the fair reward for the time and resources that a researcher invests in collecting original data, but also because it safeguards the incentive to collect original data to answer the new questions that constantly arise in the field. New quantitative (as well as qualitative) data constitute a crucial public good for the discipline. We should be careful not to undermine the incentives for producing them, while at the same time ensuring, and in some respects enhancing, the transparency and integrity of scholarly work.